

# Research<sup>2.0</sup>

*Technology Analysis for Business & Investment*

## Thought Leader Interview with Ujesh Desai, VP Marketing

**NVIDIA Corp.**

with Stephen Waite, July 28, 2010

In this Thought Leader Interview, we sit down with Ujesh Desai, VP of product marketing for NVIDIA. NVIDIA is an innovator in GPU-based computing architectures which are an important enabler of the ongoing convergence of the real and virtual worlds, or what we call RealVR. Some of the key points to emerge from our interview with Ujesh include:

- 3D-capable PCs are expected to become pervasive over the next 3-5 years.
- GPUs offer two major advantages over CPUs: Performance and Power. The performance advantage is the result of parallel processing inherent in a GPU. With respect to power, GPUs offer greater performance per watt over any other processor architecture. NVIDIA has no plans to enter the CPU market.
- NVIDIA's Fermi architecture is delivering supercomputer power to the desktop for the first time in history, with major benefits to industry (e.g., finance, oil and gas) and scientific research, as well as high-performance gaming for consumers.
- RealityServer is a revolutionary web services platform targeted for cloud computing that allows users to interact with highly complex 3D models, in life-like quality, on any web-enabled device.
- NVIDIA is at the epicenter of game development as a leader in PC graphics and a supplier to Sony PlayStation.
- In the coming decade, we'll almost certainly see CPUs and GPUs working together in supercomputers to deliver 1000x the performance of the world's fastest today.

We hope you find this interview with Ujesh Desai of NVIDIA engaging and illuminating.

*SW: I wanted to begin the interview with a question about NVIDIA's vision for 3-D computing. What are the next 3-5 years likely to bring in terms of 3-D computing technology from NVIDIA specifically?*

UD: I personally believe that in the next 3-5 years everyone will have a 3D-capable PC. Frankly there's no reason not to. The graphics processors are now powerful enough where you can get basic 3D capabilities even in a mainstream PC. Also, there's a lot more content for folks to enjoy on the PC, like games, photos, movies, and even YouTube. We've also been working with folks like The Masters and NASCAR to start streaming live sporting events in 3D and the experience is just amazing. You feel like you're there.

*SW: NVIDIA, as many people know, is a pioneer in Graphics Processing computing. What are the major advantages of GPU-based processing versus conventional CPU-based processing?*

UD: There are two main advantages of GPU processing over conventional CPU-based processing, although it is important to note that this is not a question of either/or. GPGPU, or general purpose computing on a GPU, involves the use of both GPUs and CPUs in a co-processing environment.

The first is performance. The x86 architecture has fundamentally hit a wall, and CPU performance is no longer scaling in line with the size of the computational problems they are required to solve. GPUs are a solution to this problem, as they handle data in a fundamentally different way. A good way to think about this is to take the example of a book and a program designed to count the amount of times a certain word is used in that book. Run that program on a CPU, a serial/sequential processor, and it would start on page one, read the entire book, word by word, page by page until it finished and produced a number. Give that same program to a GPU, a parallel processor, and it would tear that book into tens of thousands of pieces, and read every piece simultaneously. If you have a large computational problem that can be split into many parts, then it's likely that a GPU can process that problem orders of magnitude faster.

The next is power. GPUs offer greater performance per watt than any other processor architecture. An engineer or a scientist who used to rely on a datacenter to run his calculations can now do the same work, faster, at his desk using a workstation equipped with multiple GPUs powered by a single wall outlet. If you scale up this scenario, i.e. many large problems running across many thousands of GPUs in a datacenter style environment, then you have a solution that delivers transformative increases in speed at a fraction of the power draw. If you run a datacenter that consumes megawatts today, merely adding more CPUs and more megawatts simply is not an option. Take the example of the Nebulae supercomputer that uses NVIDIA Tesla GPUs which, last month, placed 2nd in the latest list of the Top 500 supercomputers in the world. It delivers 1.27 petaflops and consumes 2.55 megawatts. The number 1 supercomputer uses 7 megawatts to deliver 1.75 petaflops. This means Nebulae achieves DOUBLE the performance per watt by using GPUs, making it one of the most power efficient systems on the Top 500 list.

*SW: NVIDIA is helping to bring supercomputing power to the desktop for the first time in history. Tells us about the new Fermi-based Tesla<sup>®</sup> GPU; how will that technology impact corporations, businesses and consumers?*

UD: The Fermi architecture delivers a rich feature set targeted squarely at the computing space. For example, ECC — a technology that detects and corrects memory errors (a typical occurrence in large computations) — is a must-have technology in many industry segments such as finance, oil and gas, and science. Double precision accuracy is also a key feature and is typically not a feature associated with GPUs. Several detailed whitepapers have been written on the Fermi architecture, by leading industry experts, that go into detail on these and other features — they can be found here:

[http://www.NVIDIA.com/object/fermi\\_architecture.html](http://www.NVIDIA.com/object/fermi_architecture.html).

GPUs based on the Fermi architecture also bring enormous processing power to the desktop, and this has proven to be of great interest to the high performance computing community. Today, a lot of scientific research is carried out on supercomputing clusters, a shared resource that consumes hundreds of kilowatts of power and costs millions of dollars to build and maintain. As a result, researchers must compete for time on these resources, slowing their work and delaying results. The Personal Supercomputer based on NVIDIA Tesla GPUs delivers the equivalent computing power of a cluster, at 1/100th of the price and in a form factor of a standard desktop workstation. Even before we officially introduced this solution, there were researchers at institutions like MIT, Illinois, Rice and Cambridge University who were building systems themselves based on NVIDIA GPUs. They recognized the ability to get cluster performance at their desktop with these systems and increase the pace of their work by having access to their own “personal supercomputer”. The Tesla Personal Supercomputer builds on these examples.

The enormous computing capability of the Fermi architecture also brings advantages to the consumer space, especially to gamers, with features such as tessellation. Tessellation is the key feature in DX11 and it enables games to add even more complex geometry depending on the scene so that the new crop of games will get even closer to cinematic reality. While programmable shading has allowed PC games to mimic film in per-pixel effects, geometric realism has lagged behind. The most advanced PC games today use one to two million polygons per frame. By contrast, a typical frame in a computer generated film uses hundreds of millions of polygons. The Fermi architecture was built to specifically address this problem and it delivers geometry performance that is better than anything else on the market.

*SW: Let's talk about 3-D computing in the enterprise. What kinds of applications in the enterprise will be enabled by NVIDIA's GPU technology in coming months?*

UD: Digital content creation (DCC) artists, product designers, and physicians can see their ‘world’ in 3D, with perspectives that are significantly richer than traditional two-dimensional views. Physicians, for example, can now view volumetric scans like ultrasound in 3D. Businesses looking to provide large scale visualizations such as popular video walls now have a reliable, enabling technology to deliver that kind of experience. All of this is made possible with NVIDIA GPUs. This will then extend to more general business usage cases, and we also believe all displays, desktops, notebooks, tablets and handhelds will eventually become 3D capable.

*SW: How are customers responding to the AutoCAD 2011 and NVIDIA Quadro® 3-D graphics solutions?*

UD: AutoCAD 2011 adds powerful new 3D features and visual styles that enable designers, engineers and architects to create more detailed designs, and do so more quickly than before. However, to take full advantage of AutoCAD 2011, experience superior image quality, and realize true interactive performance, an NVIDIA Quadro professional graphics solution is required. Quadro lets designers dial up the complexity of their designs, visualize them in the software application, and work out any issues before moving to production, thereby increasing productivity and eliminating waste.

Each year, NVIDIA and Autodesk together spend over 2,000 engineering hours to test, certify, and optimize the latest Quadro and AutoCAD solutions. This collaborative development process increases performance and image quality, while reducing the likelihood of issues that could interrupt a customer's workflow or cause downtime.

As a result of NVIDIA's approach to developing a complete solution based on hardware, software development, and enterprise customer support, NVIDIA Quadro is the preferred choice for AutoCAD professionals.

*SW: What sets these products apart from the competition in the market?*

UD: NVIDIA Quadro is not just about a great GPU. It is that, plus much more. Quadro is a promise of high performance, high reliability, software integration and advanced capabilities. Quadro is at the core of a comprehensive ecosystem of hardware, software and services and because of this, 9 out of 10 professional workstation users rely on Quadro to get their work done. On the hardware side, there are specialized solutions for HD-SDI broadcasting, synchronization of clusters, scalable visualization solutions for power walls, and 3D stereo. On the software side, not only does NVIDIA optimize and tune drivers for the best application performance, but now more and more application developers are leveraging the advanced tools and engines NVIDIA makes available for parallel processing such as CUDA and our Application Acceleration Engines (AXE). This extensive array of solutions for artists, animators, designers, engineers and scientists is truly unrivaled.

*SW: We're seeing more virtual world applications for enterprise (e.g., Teleplace, Vennegen, Multiverse, Linden Labs). When will this real-time collaboration using virtual worlds become common in enterprise environments? What role will NVIDIA technology play?*

UD: This is a market we are currently considering but at this time we don't have anything to announce.

*SW: Tell us more about RealityServer® and the work your European-based subsidiary Mental Images is doing. How will these products and technologies impact the NVIDIA business in the next several years?*

UD: Mental Images' RealityServer is a revolutionary web services platform that lets users interact with highly complex 3D models, in life-like quality, on any web-enabled device. It's unique in that it maintains and processes its data remotely, allowing it to flexibly scale to meet application needs.

RealityServer applications leverage industry standard technologies, and virtually any interactive technology that connects to HTTP can control RealityServer – from HTML, Flash and AJAX in browsers, to Adobe AIR or PDF in documents, to apps for the Apple iPhone.

In practice, RealityServer provides professionals unprecedented freedom for interacting with complex 3D data, while bringing interactive realism to any user for the first time. We've been working privately with a number of Fortune 100 companies to implement RealityServer, and we're excited about the opportunities.

*SW: Cloud computing is a major theme in the industry and more and more server-side computing is shifting to this model. How will NVIDIA participate? Are there specific applications, companies or hosting providers that we should watch for cloud-based GPU computing?*

UD: Well, the RealityServer product we just talked about is squarely aimed at the cloud-computing space. Some of the industries we think will be able to benefit from this are in the automotive space, where product engineering teams will be able to securely share and visualize complex 3D models of cars under different lighting and environmental conditions. Architects and their clients will be able to review sophisticated architectural models, rendered in different settings, including day or night, and online shoppers will be able to

interactively design home interiors, rearrange furniture, and view how fabrics will drape, all with perfectly accurate lighting.

The RealityServer platform is comprised of an NVIDIA Tesla GPU-based server running RealityServer software from Mental Images. While photorealistic imagery has traditionally taken hours or days to create, this first-of-its-kind, integrated solution streams images of photorealistic scenes at rates approaching an interactive gaming experience.

*SW: Is there a way to settle the debate about whether NVIDIA needs to also offer a CPU? Many have said that architecturally the GPU and the CPU need to be integrated to eliminate the need to move data back and forth over what is typically a relatively slow connection.*

UD: While the CPU is a very important processor, we feel that there are already two great companies out there building CPUs, and therefore NVIDIA should focus our expertise on where we can add real value — and that's building the world's best GPUs. The combination of a great CPU and GPU is what delivers a truly amazing computing experience and this is where we'll keep our focus.

*SW: Let's switch gears and talk about NVIDIA in the consumer market. 3-D is moving into the home. NVIDIA has several 3-D products for the PC and is developing products for netbooks and tablets. Tell us about NVIDIA's vision and product strategy for 3-D in the home.*

UD: The move to 3D is really good for us. Now we have to render twice as many pixels on the screen, which means we need even more powerful GPUs! As I said earlier, I personally believe everyone will have a 3D PC in their home. I also feel this will extend to other devices like tablets and handheld mobile devices. NVIDIA recognized the importance of 3D and it's why we've invested in solutions such as our 3D Vision glasses. We're going to make sure we maintain a leadership in the 3D space, whether it's working with Hollywood on the content creation side with our Quadro products, all the way to delivering the best 3D experience in the home with our GeForce products.

*SW: 3-D gaming seems like a huge and exciting market opportunity. What's NVIDIA's role in 3-D gaming in the future?*

UD: NVIDIA is at the epicenter of game development as the leader in PC graphics and a supplier to the PlayStation. We have an extensive team focused on game developer support and are engaged in developing new standards and tools for enhancements like tessellation and 3-D stereo. At over 400 titles, we support by far the largest library of 3D games today, and we will continue to play this role of innovation within the gaming industry to bring the video game closer and closer to real life. We will also facilitate bringing 3D to multiple platforms, from TVs to PCs to Tablets to Smartphones.

*SW: The market for 3-D mobile computing seems primed to take off. Tell us about NVIDIA's vision for leading and enabling 3-D mobile computing.*

UD: The development of 3D displays is occurring at an amazing pace. Prototype displays for handheld 3D stereo displays is here today and are very compelling. Handheld gaming is taking off as cell phones become computers and their resolution improves. 3D is a logical extension of this to achieve more perceived information from a small screen. Again, games will lead this trend, and NVIDIA Tegra will provide the graphics hardware and software to lead this transition, but it will soon be followed by pictures, streaming internet video and UI development in 3D.

*SW: What is the most mind-blowing application we are likely to see with NVIDIA's GPU technology over the next year or two?*

UD: Honestly it's hard for me to answer this. The beauty of what we're doing at NVIDIA is that we've got certain ideas for what we think we can do with our GPUs, but because of the infrastructure we've put in place with CUDA, the development community is taking our GPUs and coming up with things that we never even imagined was possible. To me this is the most exciting thing about working at NVIDIA.

*SW: One last question for you, today. What kinds of advances should we expect to see in GPU computing over the five years and, if you could look out a little farther, over the next decade?*

UD: Because GPUs are enabling advances in computing to be delivered with such a broad reach, we are discovering — almost daily — new application areas that are taking advantage of GPUs. But maybe we can look at a couple of likely destinations for these advances.

In the area of high performance computing, toward the end of this decade, we'll almost certainly see CPUs and GPUs working together in supercomputers to deliver 1000x the performance of the world's fastest today. These future systems will be hundreds of times more power efficient than the world's current best.

In the area of medicine, this decade we'll see GPUs used to provide real time imaging to solve very real problems in health care. For example, our technology will make heart surgery safer by stabilizing the image of a beating heart so doctors can perform a complicated medical procedure. It will also allow more efficient computation of 3D images so that an order of magnitude less radiation is required for patient CT scans. GPUs will enable the modeling and visualization of the complex dynamics within the human body (e.g. the flow of blood through arteries and heart chambers). This will help to identify health problems in a non-invasive manner and lower the overall cost of health care. I am excited and quite proud of the impact that GPUs will contribute to science and medicine and the leadership role NVIDIA is playing in this space.

*SW: Thanks again for your time. I really enjoyed speaking with you and wish you and your colleagues at NVIDIA all the best in the future.*

#### **About Ujesh Desai:**



Ujesh Desai serves as NVIDIA's Vice President of Product Marketing, with global responsibility for campaigns, marketing and PR around the company's five product lines. These include GeForce and ION, for consumers; Quadro, for workstations; Tesla, for supercomputing; and Tegra, for mobile products. He is based in the company's Santa Clara, Calif., headquarters.

Desai joined NVIDIA in 2001 and has held a variety of roles, including general manager of desktop GPUs, which is the company's single-largest business, as well as general manager of software and multimedia products. Previously, he spent six years at Silicon Graphics, where he held a variety of management positions in the professional workstation division.

Desai received a bachelor's degree in Marketing Communications from Trenton State College, in New Jersey.

For more information about NVIDIA, go to: <http://www.NVIDIA.com/>

## **Previous Thought Leaders:**

*Think Twice: An Interview with Michael Mauboussin, Chief Investment Officer, Legg Mason Capital Management* (November 2009).

*Into the Multiverse: An Interview with Corey Bridges, Co-founder, The Multiverse Network* (December 2009).

*2D-3D: An Interview with Craig Summers, Founder and CEO, 2D-3D Video Inc.* (January 2010).

*Global Solar Technology: An Interview with Dr. Alan Rae, Technical Director* (February 2010).

*China: An Interview with Arthur Kroeber, Founder, Dragonomics Research and Advisory* (March 2010).

*Web 3D: An Interview with Anita Havele, Executive Director, Web 3D Consortium* (April 2010).

*3D Technology: An Interview with Ami Dror, Co-founder and Chief Strategy Officer, XpanD* (May 2010).

*OpenIPO: An Interview with David Marcus and James Donohue, WR Hambrecht & Co* (June 2010).

# Research<sup>2.0</sup>

## ABOUT US

Research 2.0 specializes in positioning emerging technologies and companies in a business and investment context for institutional investors. The two founders each have over 30 years of experience in applying emerging technology, company research, equity capital markets and investment management.

Our clients consist of technology companies and institutional investors that rely on in-depth, independent, thematic research and advisory services.

Feel free to visit our website ([www.research2zero.com](http://www.research2zero.com)) to learn more about us and [sign up for our free research update service](#).

Our main office is located at 1313 Washington St., #326, Boston MA 02118 U.S.A. and we have staff in New York, Connecticut and Paris, France. Our main telephone number is +1-617-848-9651.

## OUR TEAM

Kris Tuttle, Founder & CEO, [kris@research2zero.com](mailto:kris@research2zero.com)

Stephen Waite, Partner and Director of Strategy, [steve@research2zero.com](mailto:steve@research2zero.com)

Leigh Deacon, Administrative Manager, [leigh@research2zero.com](mailto:leigh@research2zero.com)

Benjamin Stein, Associate Analyst, [ben@research2zero.com](mailto:ben@research2zero.com)

Heidi Manwaring, Research Assistant, [heidi@research2zero.com](mailto:heidi@research2zero.com)

Vish Kalburgi, Associate Analyst, [vish@research2zero.com](mailto:vish@research2zero.com)

## COPYRIGHT

It is a violation of federal copyright law to reproduce or distribute any part of this publication to anyone by any means without permission. Institutional and corporate clients have internal redistribution rights and in some cases may distribute this publication externally to investors, suppliers and customers with written authorization.

## DISCLAIMER

The content of this publication is intended for informational purposes only. Research 2.0 is not a broker-dealer, investment bank or registered investment advisor. We do not make recommendations as to whether to buy or sell securities nor do we solicit orders in publicly traded securities. Proper investment decisions can only be made with a complete understanding of return expectations, risk tolerance and duration. Research 2.0 does undertake to run a "Research 2.0 Model Portfolio" which contains stocks that we publish research on. Research 2.0 employees are allowed to buy, sell and own securities in companies covered subject to certain compliance rules.